Supplementary material

No evidence of substantial underreporting of COVID-19 deaths in Taiwan during 2020.

2021-02-20

This document is the supplementary material for Chen E, Sachs M, Dickman P. No evidence of substantial underreporting of COVID-19 deaths in Taiwan during 2020. The main aim is to conduct the sensitivity analysis on using different knots in the `excessmort::excess_model`.

Set up

```
library(excessmort)
library(haven)
library(tidyr)
library(dplyr)
library(lubridate)
library(ggplot2)
```

The data of monthly deaths and population sizes were downloaded from the Department of Household Registrations, Ministry of the Interior, ROC (Taiwan) (accessed at https://ws.moi.gov.tw/001/Upload/400/relfile/0/4413/d1d7bfc0-d03a-4cb6-bac2-06353bc6d082/month/month_en.html), and the data cleaning was done in Stata, version 16.1.

```
#' Read the data
#' The -excessmort- package requires date, outcome (death), and population variables.
death_pop <- read_dta( "Data/birth_death_1981_2020_pop.dta") %>%
               mutate(date = ymd(paste(year, month, 01 , sep= ' ')),
                      outcome = death,
                      population = pop) %>%
               select(date, outcome, population)
str(death_pop)
```

```
## tibble [73 x 3] (S3: tbl_df/tbl/data.frame)
##  $ date      : Date[1:73], format: "2015-01-01" "2015-02-01" ...
##  $ outcome   : num [1:73] 14310 12122 16092 14002 12802 ...
##   ..- attr(*, "label")= chr "Number of all-cause deaths"
##   ..- attr(*, "format.stata")= chr "%10.0g"
##  $ population: num [1:73] 23440278 23445534 23449287 23452387 23456545 ...
##   ..- attr(*, "label")= chr "population size at the end opf the month"
##   ..- attr(*, "format.stata")= chr "%10.0g"
```

Model fitting

The data during January 2020 and January 2021 were first identified as `exclude_dates`, which will be excluded when computing expected counts. That is, the expected deaths from January 2020 to January 2021 will then be calculated from data of January 2015 to December 2019. The entire time length of the dataset was identified by `intervals` for further use.

```
#' Preparation for fitting the model
#' Excluded the entire 2020 and Jan 2021
#' Only 2015-2019 remained
exclude_dates <- c(seq(make_date(2020, 1, 1), make_date(2021, 01, 31), by = "day"))

intervals <- as.list(seq(as.Date("2015-01-01"), as.Date("2021-01-01"), by="months"))
```

The models were fitted by using 2, 3, 4, 6, and 8 knots per year in the spline-smoothed Poisson model to estimate excess mortality/deaths.

```
#' Fitting the model with different knots used per year.
f <- lapply(c(2,3,4,6,8), function(i){
              excess_model(death_pop,
                           exclude = exclude_dates,
                           interval = intervals,
                           start = min(death_pop$date),
                           end = max(death_pop$date),
                           knots.per.year = i,
                           verbose = FALSE,)
})
```

We re-defined a graph function entitled `excess2020` from the original `excessmort::excess_plot` to plot only January 2020 to January 2021.

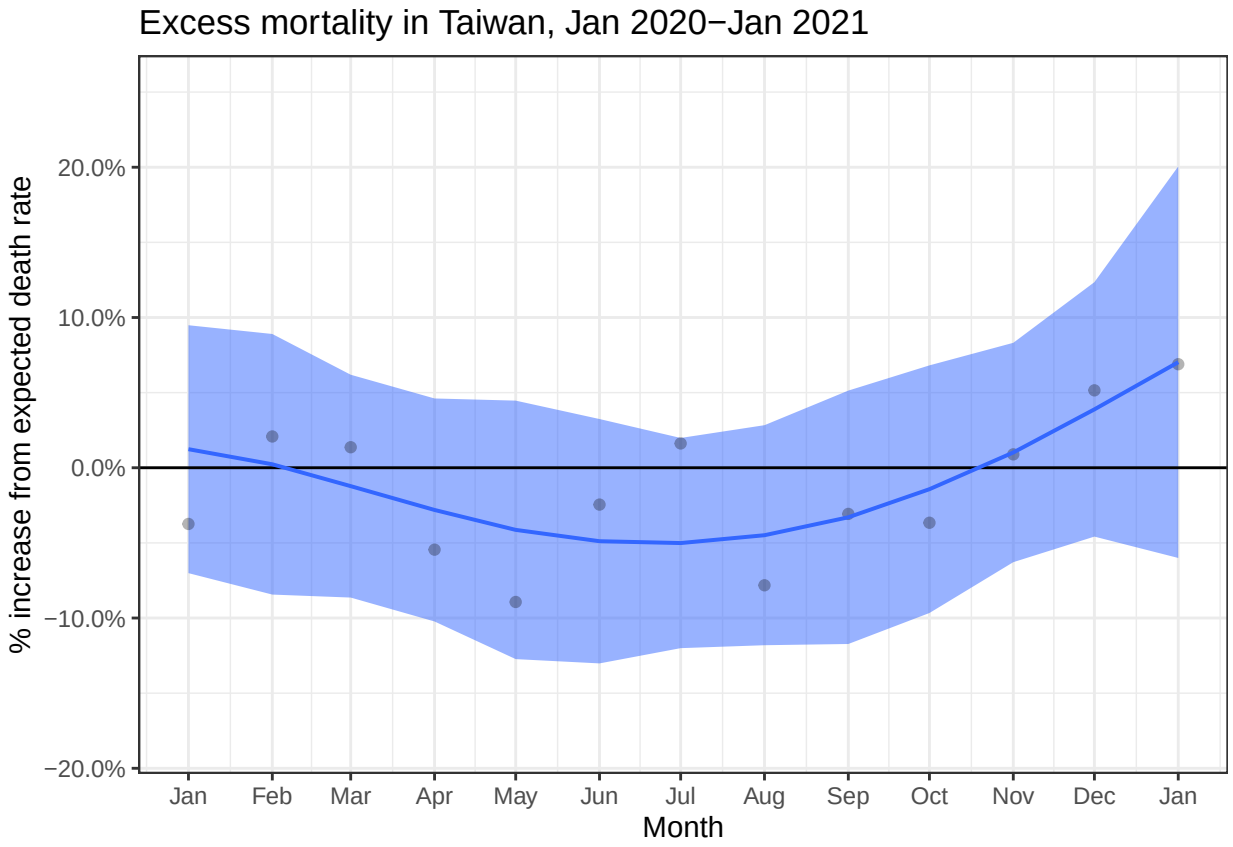
```
#'Excess mortality with 95% CI
excess2020 <-function (fit, title = "Excess mortality in Taiwan, Jan 2020-Jan 2021",
                       ylim = NULL, show.data = TRUE, alpha = 0.05)
  {requireNamespace("ggplot2")
  z <- qnorm(1 - alpha/2)
  p <- with(fit, data.frame(date = date, y = (observed - expected)/expected,
                            increase = fitted, sd = sd, se = se)) %>%
    ggplot(aes(date, y)) +
    geom_ribbon(aes(ymin = increase - z * se, ymax = increase +  z * se),
                alpha = 0.5, fill = "#3366FF") + geom_hline(yintercept = 0) +
    ylab("% increase from expected death rate") +
    scale_y_continuous(labels = scales::percent) +
    scale_x_date(date_breaks = "1 month", date_labels = "%b",
                 limits = c(as.Date("2020-01-01"), as.Date("2021-01-01") )) +
    ggtitle(title) +
    xlab("Month") + theme_bw()
  if (show.data)
    p <- p + geom_point(alpha = 0.3)
  if (!is.null(ylim))
    p <- p + coord_cartesian(ylim = ylim)
  return(p + geom_line(aes(y = increase), size = 0.7, col = "#3366FF"))
  }
```

Three knots per year were used in the spline-smoothed Poisson model in the article to estimate excess mortality/deaths.

```
#'The model used in the original article had 3 knots per year.
excess2020(f[[2]])
```

```
## Warning: Removed 60 rows containing missing values (geom_point).
```

```
## Warning: Removed 60 row(s) containing missing values (geom_path).
```



Sensitivity analysis

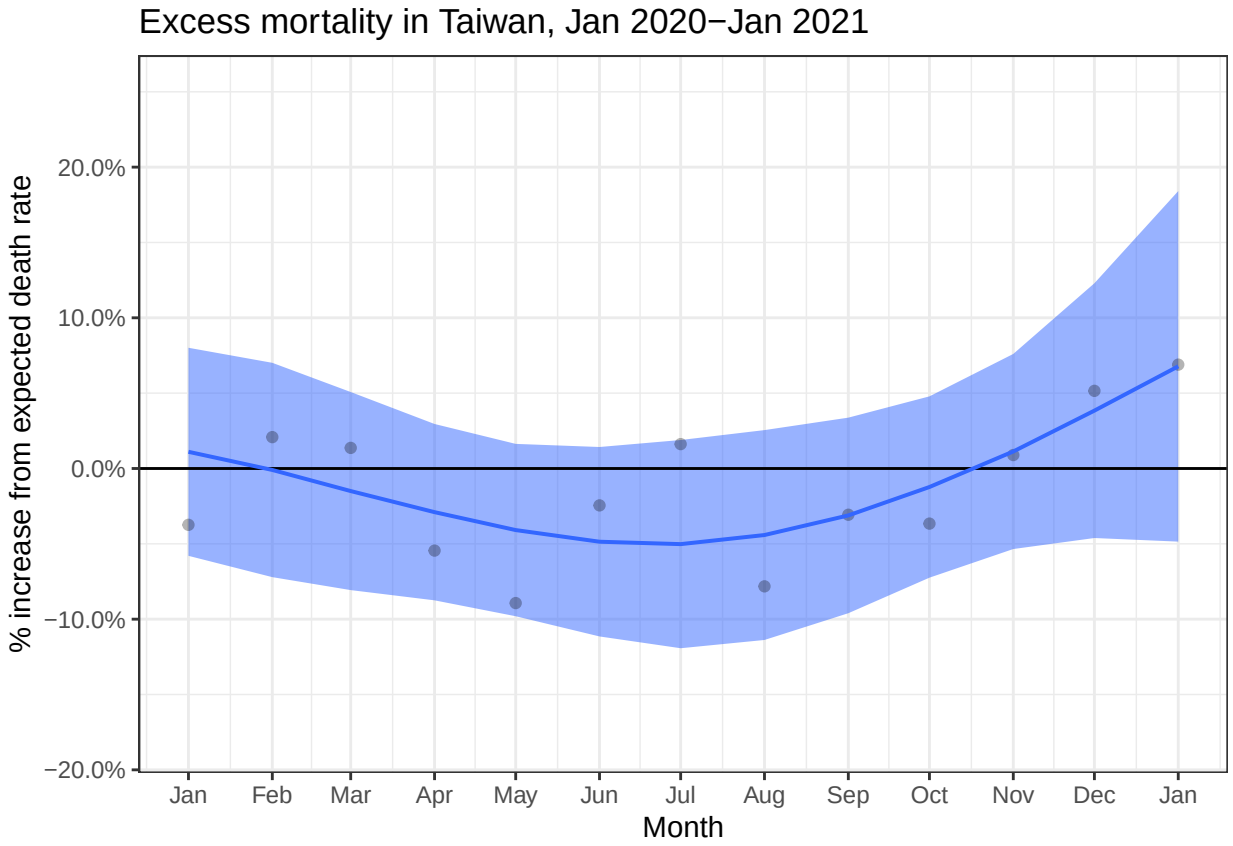
Sensitivity analysis was conducted by using different numbers (2, 4, 6, and 8) of spline knots used per year in the `excessmort::excess_model`.

2 knots per year

```
excess2020(f[[1]])
```

```
## Warning: Removed 60 rows containing missing values (geom_point).
```

```
## Warning: Removed 60 row(s) containing missing values (geom_path).
```

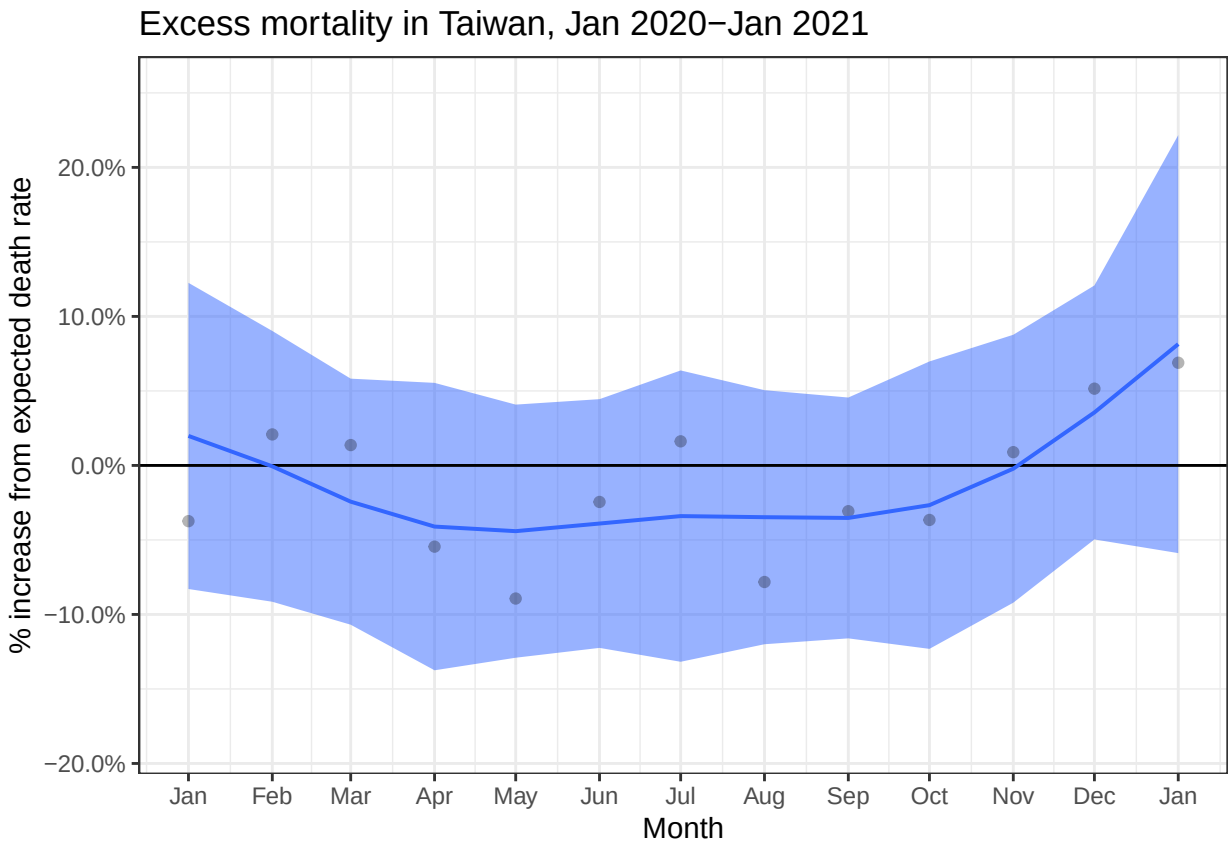


4 knots per year

```
excess2020(f[[3]])
```

```
## Warning: Removed 60 rows containing missing values (geom_point).
```

```
## Warning: Removed 60 row(s) containing missing values (geom_path).
```

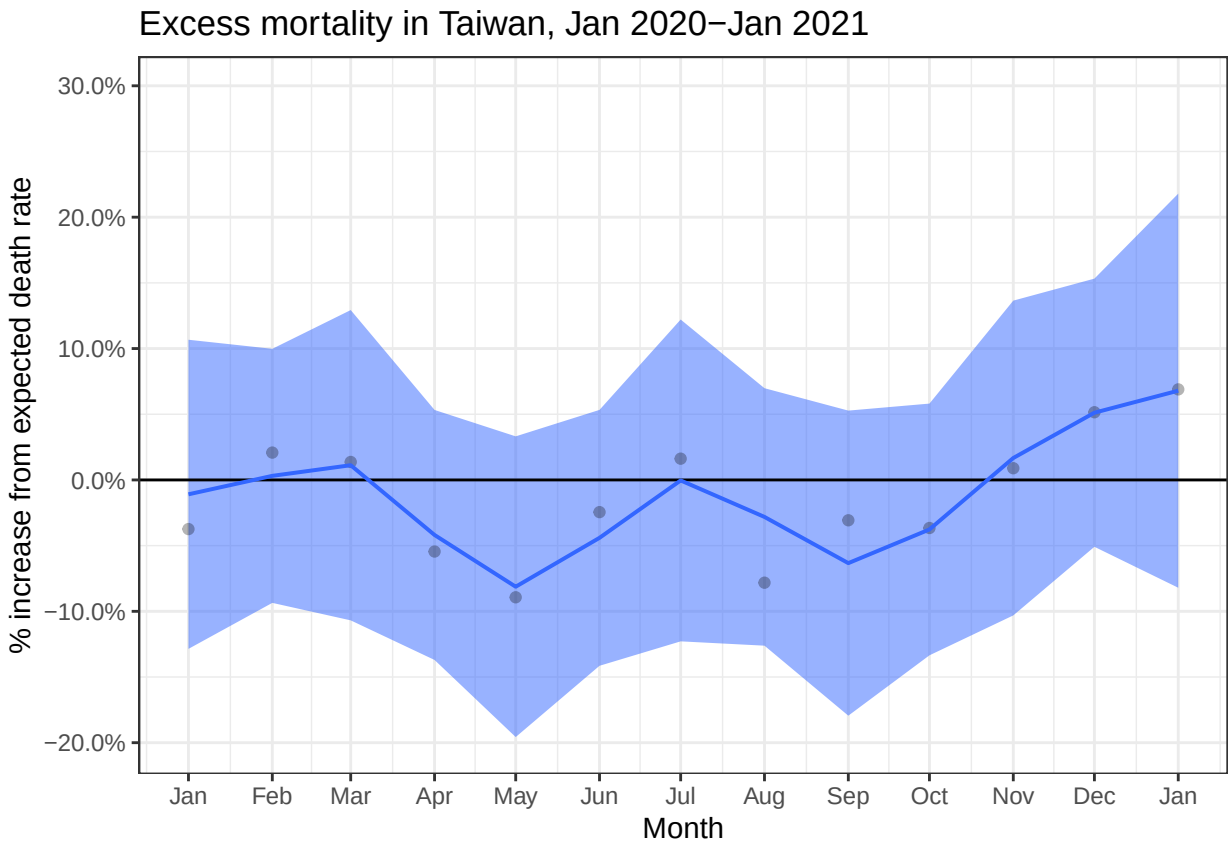


6 knots per year

```
excess2020(f[[4]])
```

```
## Warning: Removed 60 rows containing missing values (geom_point).
```

```
## Warning: Removed 60 row(s) containing missing values (geom_path).
```



8 knots per year

```
excess2020(f[[5]])
```

```
## Warning: Removed 60 rows containing missing values (geom_point).
```

```
## Warning: Removed 60 row(s) containing missing values (geom_path).
```